An Open Resource for Collaborative Biomedical Big Data Training

The University of California San Diego

PI: Rommie E. Amaro and Ilkay Altintas de Callafon Grant Number: 1R25GM114821-01

Of all the resources required to make gaining insight from big data a success, perhaps the most important is the human one. A major challenge to the big data community generally and especially, the biomedical big data community is training and education of the current and next generation of biomedical scientists. We must work collectively to address this critical challenge. What we seek to do through this proposed project is maximize the impact of biomedical big data training through a large-scale collaborative approach, and to create a training and education framework for other educators (a.k.a., teachers, instructors) and/or learners (a.k.a., students, trainees, researchers) that enables them to construct and deliver customized modules or courses that deliver the highest value to their particular application. Our vision is to cultivate a high-quality, well-informed, freely accessible knowledge and data community effort around training and education in biomedical big data research, through the Biomedical Big Data Training Collaborative (BBDTC). The end-to-end BBDTC open online training framework is a repository allowing faculty, researchers and students access to a state-of-the-art training model over the years to come. To develop such an environment, we employ current best practices and build upon our existing efforts. Initially we focus on building the BBDTC along with example courses, lecture content and hands-on application use cases for biomedical big data training. We will also communicate best practices for developing course content and delivering it to a wide-range of trainees along with associated adaptive learning approaches and assessments. In addition, we will deliver customizable virtual machines (VMs) including the course materials, hands-on tools and example data and additional assessment and make sure that these VMs are portable to a variety of environments. Specific aims in the project include development of: (1) Biomedical Big Data Curriculum; (2) Biomedical Big Data MOOC Framework; (3) Biomedical Big Data Tool Box; and (4) Repository Interfaces to Engage Community Stakeholders. The significance of our approach is that the BBDTC will enable the development of many more courses and training modules (whether they are full-scale MOOCs or much smaller, more targeted units). Although we focus on the present "mission critical" challenges defined by the NIH and biomedical community, we build the BBDTC framework in a way that will allow it to evolve over the years, not just by one person but by a community of biomedical big data researchers as a collective force to handle training challenges of the future. PUBLIC HEALTH RELEVANCE: Our vision is to cultivate a high-quality, well-informed, freely accessible knowledge and data community effort around training and education in biomedical big data research, through the Biomedical Big Data Training Collaborative (BBDTC). The BBDTC open online training framework will enable faculty, researchers and students access to a state-of-the-art biomedical big data training model over the years to come.